# Big Data and Cloud Computing

**Mrs. Premalatha P, Mrs. Marrynal S. Eastaff**

***Abstract*— Cloud computing is one of the most significant shifts in modern ICT and service for enterprise applications and has become a powerful architecture to perform large-scale and complex computing. Big data provides users the ability to use commodity computing to process distributed queries across multiple datasets and return resultant sets in a timely manner. Big data utilizes distributed storage technology based on cloud computing rather than local storage attached to a computer or electronic device. Big data evaluation is driven by fast-growing cloud-based applications developed using various categories of big data. Cloud computing, big data and its applications, advantages are likely to represent the most promising new frontiers in science. Clouds are also being used to deal with the Big data to effectively store and exploit the unstructured data of the organizations. This paper presents an overview of the cloud computing scenario today, different examples of the cloud services, different enterprises in the field of cloud computing are being mentioned in the paper . How cloud is related with big data and what are the possible solutions of big data in today's scenario is also discussed in the paper.**

**Index Terms— Big data, Cloud computing, Providers.**

## I. INTRODUCTION

Cloud Computing is defined as a collection of integrated and networked hardware, software and Internet infrastructure called a platform i.e. using the Internet for communication and transporting hardware, software and networking services to clients [11]. This platform hides the complexity and details of the underlying infrastructure from users and applications by providing very simple graphical interface or API (Applications Programming Interface) and also provides on-demand services that are always on, anywhere, anytime and anyplace. But as more and more information are placed in the cloud, concerns begin to grow about the security of the cloud environment.

Big data is known as a datasets with size beyond the ability of the software tools that used today to manage and process the data within a dedicated time. With Variety, Volume, Velocity Big Data such military data or other unauthorized data need to be protected in a scalable and efficient way [1]. Information privacy and security is one of most concerned issues for Cloud Computing due to its open environment with very limited user side control [2].

Cloud computing is a way to increase the capacity or add capabilities dynamically without investing in new infrastructure, training new personnel, or licensing new software. The cloud helps organizations and enables rapid on demand provisioning of server resources such as CPUs, manage storage, bandwidth, and share and analyze their Big Data in a reasonable way.

**Mrs. Premalatha P,** Asst Professor, Hindusthan College of Arts and Science, Coimbatore, India.

**Mrs, Marrynal S. Eastaff,** Asst Professor, Hindusthan College of Arts and Science, Coimbatore, India

## II. CLOUD COMPUTING AND DELIVERY MODELS

Cloud computing revolutionizes the way information is handled, the typical deployment models for cloud computing includes: infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS) and hardware as a service (HaaS).

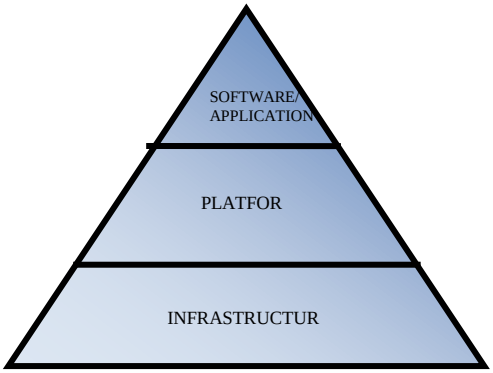


Fig. 1. Cloud Computing Service delivery models

### 2.1. Infrastructure as a Service (IaaS):

In this model the consumers are given full freedom to manage their data on the server. Here the service provider is only responsible for raw storage, computing power, networks, firewalls, and load balancers and this is often manifested as a virtual machine. Benefits of IaaS include increased financial flexibility, choice of services, business agility, cost- effective scalability, and increased security. Infrastructure as a Service (IaaS), e.g. virtual machines, networks, storage, or servers, is the most basic building block and includes anything (real or virtual) you would expect inside a data center [3].

### 2.2. Platform as a service (PaaS):

Platform as a Service is a level above Infrastructure as a service (IaaS). In the PaaS model, consumers are provided with an operating system, programming language execution environment, database, and web server. They are not concerned with the cost and management in the hardware and software layers. PaaS is the use of cloud computing to provide platforms for the development and use of custom applications [4]. The advantages of using PaaS includes: lowering risks by using pretested technologies, promoting shared services, improving software security, and lowering skill requirements needed for new systems development [3]. As related to big data, PaaS provides companies a platform for developing and using custom applications needed to analyze large quantity of unstructured data at a low cost and low risk in a secure environment. Therefore maintaining the integrity of

applications and enforcing accurate authentication checks during the transfer of data across the entire networking channels is fundamental.

### 2.3. Software as a Service:

Software as a service (SaaS) is the level above Platform as a service (PaaS). In this model, consumers are given access only to the application software, which can be run remotely from the data centers of the cloud service provider. The provider is responsible for the maintenance and support of the infrastructure and operating platforms. The main advantage of SaaS is that this solution allows businesses to shift the risks associated with software acquisition while moving IT from being reactive to proactive [9]. Benefits of using SaaS include: easier software administration, automatic updates and patch management, software compatibility across the business, for the time and number of users necessary.

### 2.4. Hardware as a service (HaaS):

HaaS offers only the hardware. HaaS serves the following purposes in managed services:

- Involves a contract for the maintenance and administration of hardware systems. This type of service may be remote or on site, depending on the hardware setup requirements.
- Helps users manage hardware licensing requirements.

In collective computing environments, HaaS participants often use Internet Protocol (IP) connections to utilize the computing power of remote hardwaee. A user sends data to a provider, and the provider's hardware performs necessary actions to the data and then sends back the results. These types of agreements help individual businesses lease computing power, rather than invest in additional on-site hardware. Some of the most popular types of HaaS models are classified as cloud computing services, in which data storage media and even active computing hardware are components of a remotely provisioned service for users [12].

## III. Types of Clouds

The Cloud Computing model has three types of clouds model – the public cloud, the private cloud, and the hybrid cloud.

### 3.1. Public cloud:

A public cloud is one based on the standard cloud computing model, in which a service provider makes resources, such as applications and storage, available to the general public over the Internet. Public cloud services may be free or offered on a pay-per-usage model [5].


Fig.2 Public Cloud

### 3.2. Private cloud:

Private cloud is a type of cloud computing that delivers similar advantages to public cloud, including scalability and self-service, but through a proprietary architecture. Unlike public clouds, which deliver services to multiple organizations, a private cloud is dedicated to a single organization [5].

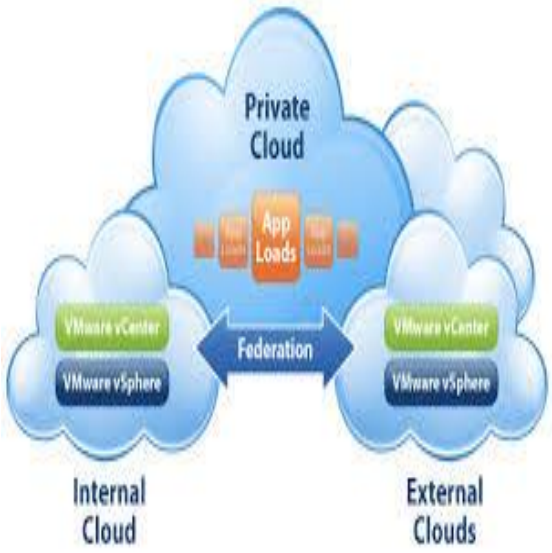


Fig.3 Private Cloud

### 3.3. Hybrid cloud:

Hybrid cloud is a cloud computing environment which uses a mix of on-premises, private cloud and public cloud services with orchestration between the two platforms. By allowing workloads to move between private and public clouds as computing needs and costs change, hybrid cloud gives businesses greater flexibility and more data deployment options [6].

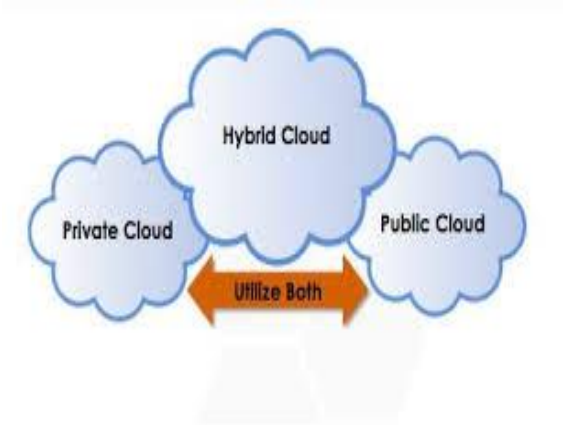


Fig.4 Hybrid Cloud

### 3.4. Community cloud:-

Community cloud is a private cloud that is shared by several customers with similar security concerns and the same data and applications sensitivity.

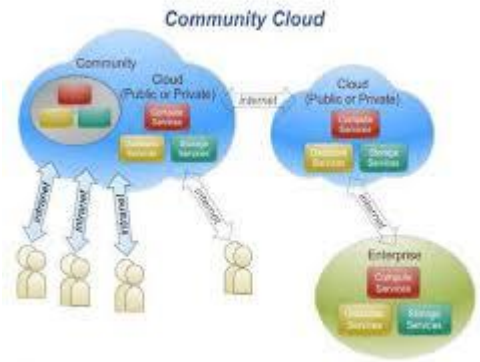


Fig. 5 Community Cloud

## IV. Big Data And The Cloud

Cloud computing has managed to make the world's already open for data storage even more voracious. When it comes to storage, everything is getting bigger, whether it's an individual disk, a storage system or a cloud-based repository. In many traditional cases, redundancy is achieved by replicating data from primary storage devices to target arrays at the data center or an off-site location. Mirroring data in that way provides protection but also consumes lots of storage, particularly when organizations make multiple copies of data for greater redundancy. The approach becomes particularly unwieldy for organizations that deal with petabytes or more of data.

Cloud computing provides vast computing resources on demand. It has become important due to the growth of "big data": the large, complex, datasets now being created in almost all fields of activity, from healthcare to e -commerce.

Horizontal scaling refers to the ability to replace a single small computing resource with a bigger one to account for increased demand. Cloud computing supports this by making various resource types available to switch between them.

Vertical scaling achieves elasticity by adding additional instances with each of them serving a part of the demand. Software like Hadoop is specifically designed as distributed systems to take advantage of vertical scaling.

Big Data is a data analysis methodology enabled by recent advances in technologies and architecture which support high velocity data capture, storage, and analysis. However, big data entails a huge commitment of hardware and processing resources, making adoption costs of big data technology prohibitive to small and medium sized businesses. The characteristics of big data present data storage and data analysis challenges to businesses. Analyzing big data is done using a programming paradigm called MapReduce. The MapReduce paradigm requires that huge amounts of data be analyzed. Cloud computing offers the promise of big data implementation to small and medium sized businesses. Data storage using cloud computing is a viable option for small to medium sized businesses considering the use of Big Data analytic techniques. Cloud computing is on-demand network access to computing resources which are often provided by an outside entity and require little management effort by the business.

## V. Big Data Cloud Providers

Some of the cloud providers that offer IaaS services that can be used for big data include Amazon.com, AT&T, GoGrid, Joyent, Rackspace, IBM, and Verizon/Terremark. Currently, one of the most high-profile IaaS service providers is Amazon web Services with its Elastic Compute Cloud (Amazon EC2). Amazon EC2 offers scalability under the user's control, with the user paying for resources by the hour. Amazon also offers other big data services to customers of its Amazon web Services portfolio. These include the following [5]:

### 4.1 Amazon Elastic MapReduce:

This is targeted for processing huge volumes of data. Elastic MapReduce utilizes a hosted Hadoop framework running on EC2 and Amazon Simple Storage Service (Amazon S3). Users can now run HBase.

### 4.2 Amazon DynamoDB:

A fully managed not only SQL (NoSQL) database service. DynamoDB is a fault tolerant, highly available data storage service offering selfprovisioning, transparent scalability, and simple administration. It is implemented on SSDs (solid state disks) for greater reliability and high performance.

### 4.3 Amazon Simple Storage Service (S3):

A web-scale service designed to store any amount of data. The strength of its design center is performance and scalability, so it is not as feature laden as other data stores. Data is stored in "buckets" and you can select one or more global regions for physical storage to address latency or regulatory needs.

### 4.4 Amazon High Performance Computing:

Tuned for specialized tasks, this service provides low-latency tuned high performance computing clusters. Most often used by scientists and academics, HPC is entering the mainstream because of the offering of Amazon and other HPC providers. Amazon HPC clusters are purpose built for specific workloads and can be reconfigured easily for new tasks.

### 4.5 Amazon RedShift:

Available in limited preview, RedShift is a petabyte-scale data warehousing service built on a scalable MPP architecture. Managed by Amazon, it offers a secure, reliable alternative to in-house data warehouses and is compatible with several popular business intelligence tools.

## VI. Big Data Cloud Storage

The cloud storage challenges in big data analytics fall into two categories: capacity and performance. Scaling capacity, from a platform perspective, is something all cloud providers need to watch closely. Data retention continues to double and triple year-over-year because customers are keeping more of it. Cloud storage is effectively a boundless data sink. Importantly for computing performances is that many solutions also scale horizontally, i.e. when data is copied in parallel by cluster or parallel computing processes the throughput scales linear with the number of nodes reading or writing [8].

## VII. Conclusion

Benefits of cloud computing are tempered by two major concerns – security and loss of control. The researchers are focusing their efforts on how to manage, handle and process the huge amount of data known as Big data which deals with three concepts Volume, Variety and Velocity. Managing and processing of big data have many problems and require more efforts to handle these requirements. Security is one of the challenges that arise when systems try to handle the concept of big data. Furthermore, the key issues in big data in clouds were highlighted. The size of data at present is huge and continues to increase every day. The variety of data being generated is also expanding. The velocity of data generation and growth is increasing because of the proliferation of mobile devices and other device sensors connected to the Internet. These data provide opportunities that allow businesses across all industries to gain real-time business insights. The use of cloud services to store, process, and analyze data has been available for some time; it has changed the context of information technology and has turned the promises of the on-demand service model into reality. In this study, we presented a review on the rise of big data in cloud computing.

## References

[1] Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G…Zaharia, M. (2010, April). A view of cloud computing. Communications of the ACM, 53(4), 50-58. DOI: 10.1145/1721654.1721672.

[2] Intel IT center, "Peer Research Big Data Analytics ", Intel's IT Manager Survey on How Organizations Are Using Big Data, AUGUST 2012.

[3] International Journal of Computer Science and Software Engineering (IJCSSE), Volume 3, Issue 1, December 2014 ISSN (Online): 2409-4285 www.IJCSSE.org Page: 78-85 -"A Survey of Big Data Cloud Computing Security"

[4] http://www.informationweek.com/big-data/bigdataanalytics/big-databrings-big-security-problems/d/did/1252747.

[5] http://www.semantikoz.com/blog/big-data-as-a-service-definition-classification/

[6] https://www.google.co.in/?gws_rd=ssl#q=private+cloud+computing

[7] https://www.google.co.in/?gws_rd=ssl#q=hybrid+cloud+computing

[8] Storing Big Data- The Rise of the Storage Cloud, Young-Sae Song, and December 2012.

[9] http://data-informed.com/cloud-computing-experts-detailbig-data-security-and-privacy-risks/

[10] http://www.vormetric.com/data-securitysolutions/applications/big-data-security

[11] http://www.aseit.com.au/solutions/cloud-overview/

[12] https://www.techopedia.com/definition/13965/hardware-as-a-service-haas